

Network Models of Route Choice

Benjamin G. Heydecker CMath FIMA, Centre for Transport Studies, UCL

1 Introduction

Network models are used in transport studies to explore the effects of individual travellers' route choice. These model the likely consequences of changes in the demand for travel, facilities provided, and ways in which demand is assigned. This enables planners to anticipate the response of travellers to changes and developments when investigating their effects on network performance. The outputs calculated from these models include estimates of various costs and flows.

Here, we consider how route choice models can be formulated, analysed and solved using mathematical methodologies. This illustrates how models of traveller behaviour can be used to anticipate response to changes and interventions, and so to support the design of systems and their use that lead to beneficial provision and management. We examine the difference between network performance when individual travellers choose their own routes and when they are assigned optimally. A simple example network is used to illustrate the effects of various management measures.

2 Route choice and other traveller choices

2.1 Demand for travel

Personal travel arises from a need to attend at another location in the near future. In traffic management, an objective is to design and manage transport systems so that they serve the travelling public well after individuals have exercised their various choices. Here we consider route choice during an interval short enough that the resulting flows are approximately constant but long enough that any transients from variation in flow are not influential. We therefore consider the demand for travel in the network as a fixed flow T_{od} from each origin o to each destination d , and represent the collection of these flows as T . The model represents interactions between these different demands for travel through network effects.

2.2 Transport networks

Transport networks in which we consider route choice include the strategic road network of England [1, 2] – see Figure 1. Certain points are identified within these networks to represent the origins and destinations of journeys made from surrounding zones in the region. Numbers of these origins and destinations increase approximately linearly with the spatial area of the region and hence like the square of its diameter. Numbers of origin–destination pairs available for travel increase like the product of the numbers of origins and destinations, and hence like the fourth power of the diameter. In a typical transport network, many origin–destination pairs are connected by several routes, usually providing a reasonable choice.

The network itself can be described as a set of nodes and a set of directed links that each connect a certain pair of nodes. Each node represents a terrestrial location, and each link represents an opportunity for travel provided by a transport corridor. Links are defined by their start and end nodes, with various additional attributes including spatial length, travel time in free-flow

conditions and capacity, but also possibilities such as attractiveness and price charged.



Figure 1: Map of the strategic road network in England.

An important feature of transport networks used for personal travel is that each link can be used by several different routes. This can include routes that serve the same origin–destination pair, and also routes that serve distinct origin–destination pairs, which can lead to interaction between their respective costs of travel and flows. Individual travellers choose routes through the network between their origin and destination, and hence contribute directly to the route flows, but the travel conditions that they experience depend on the total flow on the links that they use. For this reason, models of networks require explicit representation of the flows and costs both on routes and on links.

The relationship between link and route flows can be expressed as

$$v = \phi t \quad (1)$$

where v is a vector of link flows with v_a the flow on link a , $\forall a \in L$, L is the set of links in the network, and ϕ is a binary link-route incidence matrix with

$$\phi_{ap} = \begin{cases} 1, & \text{if route } p \text{ includes link } a, \\ 0, & \text{otherwise,} \end{cases} \quad \forall a \in L, \forall p \in P_{od}, \forall od.$$

P_{od} is the set of routes from origin o to destination d , $\forall od$, and t is a vector of route flows with t_p the flow on route p .

The analysis of zone and link numbers shows that there are usually many more routes through a network than there are links. In such cases, the link-route incidence matrix ϕ cannot have full

column rank, with the consequence that several distinct route flow vectors can give rise to the same link flow vector. Bar Gera [3] for example reports practical networks with over 10^3 trip end zones, 10^4 nodes and links, 10^6 origin–destination pairs and 10^9 routes used.

2.3 Route choice modelling

To model travellers' choices of routes, we require a principle that represents their behaviour. An important example of this is the *equilibrium principle of route choice* that has been adopted widely and analysed extensively since it was presented by Wardrop [4] in 1952:

The journey times on all routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route. (W1)

This includes two simplifications to be recognised when interpreting its application. First, there is a tacit assumption that all travellers select their routes according to the criterion of minimal journey time: this is usually extended to a measure of generalised cost calculated as a weighted sum of components such as travel time, distance and price charged. The second simplification is that all travellers are supposed to be accurately informed about the prospective cost of travel on all available routes before they depart.

Here we consider a model of homogeneous fully informed travellers, which provides a basis for further development. In this case, the equilibrium condition (W1) can be expressed as

$$\begin{aligned} t &\geq 0 \\ t^T \cdot [C - \theta K] &= 0 \\ C - \theta K &\geq 0 \end{aligned} \quad (2)$$

where C is a vector of route costs, with C_p the cost of travel on route $p \in P_{od}$, $\forall od$, K is a vector of minimum cost of travel, with K_{od} the minimum cost between origin o and destination d , $\forall od$, and θ is a binary matrix that identifies the origin–destination pair served by each route:

$$\theta_{p\ od} = \begin{cases} 1, & \text{if } p \in P_{od}, \\ 0, & \text{otherwise,} \end{cases} \quad \forall od.$$

2.4 Cost of travel

Because travel conditions on a link depend on all the traffic that it carries, the cost of travel along links depends on the sum of the route flows using them. This dependence is represented as cost function $c(v)$, where component $c_a(v)$ represents the cost of travel on link a , which might depend on the flows on several links of the network. We suppose throughout that these functions are continuously differentiable. A plausible requirement is that the cost of travel on each link increases with the flow on that link:

$$\frac{\partial c_a}{\partial v_a} > 0, \quad \forall a \in L,$$

in which case elements on the principal diagonal of the Jacobian matrix $J = \nabla_v c$ are strictly positive. A further requirement is that of *separability*: the cost of travel on each link depends only on the flow on that link:

$$\frac{\partial c_a}{\partial v_b} = 0, \quad \forall a \in L, \forall b \in L \setminus \{a\},$$

in which case J is strictly diagonal.

These requirements can be generalised to a function $c(\cdot)$, $\Re^m \rightarrow \Re^m$, that is *monotone* if

$$[c(t) - c(s)]^T \cdot (t - s) \geq 0 \quad (3)$$

and is *strictly monotone* if equality arises only when $t = s$. Any separable function that is (strictly) increasing is (respectively strictly) monotone, whilst monotonicity allows for some interactions between costs on some links and flows on others. Here, we will suppose that the link cost functions are strictly monotone.

A class of functions that is widely adopted for road traffic assignment was proposed by the US Bureau of Public Roads [5], which are given as a constant plus a variable component of travel time that is a power function of flow:

$$c_a(v_a) = b_a + \alpha \left(\frac{v_a}{Q_a} \right)^\beta t_a, \quad \forall a \in L, \quad (4)$$

where b_a is the travel cost when $v_a = 0$, Q_a is a reference flow, t_a is the free-flow travel time, and α and β are strictly positive parameters.

The reference flow Q_a is often interpreted as the capacity of link a , in which case αt_a is the increase in travel time above free-flow when flow is equal to capacity. The parameter β controls the increase in cost as flow increases. These parameters are often set so that $\alpha = 0.15b_a/t_a$ and $\beta = 4$, though these can be varied to represent different relationships and can be specified for links of the network individually. The value $\beta = 1$ corresponds to affine link costs that increase with flow at rate $\alpha t_a/Q_a$. These functions are separable and strictly increasing and hence strictly monotone.

The cost of travel on routes can be calculated as the sum of those on the links that they include. Thus

$$C = \phi^T c(v) = \phi^T c[\phi(t)] \quad (5)$$

where ϕ is the same link-route incidence matrix as is used in (1).

3 Traffic assignment

3.1 Introduction

According to Wardrop's equilibrium principle (W1), the flows on the routes through a network depend on the costs incurred on them. If costs were independent of flow, then route choice would be as well and unless there is exact equality on minimum cost routes, all flows between each origin–destination pair would be on a single route. However, in practice the limited capacity of the links of a network leads to costs that depend on flows. Thus:

Route flows t_p	\rightarrow	link flows v_a	(1)
Link flows v_a	\rightarrow	link costs c_a	(e.g. 4)
Link costs c_a	\rightarrow	route costs C_p	(5)
Route costs C_p	\rightarrow	route flows t_p	(2).

This results in adjustment of route flows to balance route costs leading to several routes being used at the same cost between each origin–destination pair. Resolving and analysing these interactions between costs and flows is the topic of traffic assignment, which we consider here.

3.2 Assigning the demand

For route flows \mathbf{t} to be a feasible assignment of demand \mathbf{T} , we require that they are positive and sum to the required values:

$$\mathbf{t} \geq 0, \quad \text{and} \quad \boldsymbol{\theta}^T \mathbf{t} = \mathbf{T} \quad (6)$$

where $\boldsymbol{\theta}$ is the same route – origin–destination incidence matrix as used in (2).

We refer to the set of feasible route flows as \mathcal{D} ; according to the feasibility requirements (6), \mathcal{D} is convex, closed and bounded, hence compact.

An equivalent expression to (2) for $\mathbf{t}^* \in \mathcal{D}$ to be an equilibrium, which we shall develop here, is

$$-\mathbf{C}(\mathbf{t}^*)^T \cdot (\mathbf{t} - \mathbf{t}^*) \leq 0, \quad \forall \mathbf{t} \in \mathcal{D}, \quad (7r)$$

where $\mathbf{C}(\mathbf{t})$ is the vector of route costs when the route flows are \mathbf{t} . This identifies whether or not an assignment is an equilibrium. We will develop it to provide a practical test for equilibrium and solution procedures.

Using (1) and (5) we can deduce from (7r) the link-based expression for $\mathbf{v}^* = \boldsymbol{\phi} \mathbf{t}^*$ to be an equilibrium:

$$\begin{aligned} -\mathbf{C}(\mathbf{t}^*)^T \cdot (\mathbf{t} - \mathbf{t}^*) &\leq 0, \quad \forall \mathbf{t} \in \mathcal{D}, \quad (7r) \\ \Leftrightarrow -\mathbf{c}(\mathbf{v}^*)^T \boldsymbol{\phi} \cdot (\mathbf{t} - \mathbf{t}^*) &\leq 0, \quad \forall \mathbf{t} \in \mathcal{D}, \quad (\text{from } 5) \\ \Leftrightarrow -\mathbf{c}(\mathbf{v}^*)^T \cdot (\mathbf{v} - \mathbf{v}^*) &\leq 0, \quad \forall \mathbf{t} \in \mathcal{D}, \quad (\text{from } 1), \quad (7l) \end{aligned}$$

where $\mathbf{v} = \boldsymbol{\phi} \mathbf{t}$. This link-based variational inequality formulation has the advantage of being defined in a space of dimensionality – number of links rather than number of routes.

This variational inequality formulation of equilibrium was discovered by Smith [6]. It tests an assignment as a candidate for equilibrium by considering costs only at that assignment. According to this expression, in equilibrium for each origin–destination pair the route that currently has least cost of travel has the same cost as any that is used, which provides a convenient test for equilibrium and also the basis for solution methods.

The quantity $S = \mathbf{C}(\mathbf{t})^T \cdot \mathbf{t} = \mathbf{c}(\mathbf{v})^T \cdot \mathbf{v}$ is the total cost incurred by assignment \mathbf{t} and the corresponding link flows $\mathbf{v}(\mathbf{t})$ provides a measure of performance of the network with these flows. It also provides the interpretation of the condition (7) that $\mathbf{t}^* \in \mathcal{D}$ is an equilibrium assignment if any feasible reassignment of traffic at current costs would increase the total cost of travel. This will occur exactly when flow is reassigned to a route that currently has greater cost.

3.3 Solution approaches

Procedures to calculate assignments are important for the transport planning industry. The computational demand associated with this can be substantial, so efficient methods are important. A substantial literature has been established [3, 7, 8] and work on this continues. Here, we present a simple approach that generates convergent sequences and hence establishes the existence of equilibrium assignments as solutions to (7).

Starting from the route-based formulation (7r) of equilibrium, we consider the programme for $\mathbf{t} \in \mathcal{D}$:

$$g(\mathbf{t}) = \max_{\mathbf{s} \in \mathcal{D}} -\mathbf{C}(\mathbf{t})^T \cdot (\mathbf{s} - \mathbf{t}), \quad (8)$$

which is known as the *gap* at assignment \mathbf{t} . Setting $\mathbf{s} = \mathbf{t}$ yields 0 so $g(\mathbf{t}) \geq 0$, whilst according to condition (7r), at equilibrium flows $\mathbf{t} = \mathbf{t}^*$ the value of the maximand in (8) is negative for all feasible assignments \mathbf{s} so that $g(\mathbf{t}^*) = 0$. The gap $g(\mathbf{t})$ therefore provides a measure of the departure of \mathbf{t} from equilibrium.

The solution to (8) can be calculated conveniently without testing each feasible assignment $\mathbf{s} \in \mathcal{D}$. Because \mathbf{t} is exogenous to the programme (8), the solution is $\mathbf{s}^* \in \mathcal{D}$ that solves the linear programme

$$\min_{\mathbf{s} \in \mathcal{D}} \mathbf{C}(\mathbf{t})^T \cdot \mathbf{s}.$$

This is achieved by for each origin–destination pair od assigning all of the demand T_{od} to the route $p \in P_{od}$ that minimises $C_p(\mathbf{t})$: we denote this assignment as $\mathbf{s}^*(\mathbf{t})$. Because costs $\mathbf{C}(\mathbf{t})$ are continuous in flows \mathbf{t} , so is $\mathbf{C}(\mathbf{t})^T \cdot \mathbf{s}^*(\mathbf{t})$ and hence is $g(\mathbf{t})$.

At each $\mathbf{t} \in \mathcal{D}$ we consider two cases according to the value of $g(\mathbf{t})$:

1. $g(\mathbf{t}) = 0$.

The least cost of travel $\min_{p \in P_{od}} C_p(\mathbf{t})$ for each origin–destination pair od is no less than the current cost of travel: \mathbf{t} is an equilibrium.

2. $g(\mathbf{t}) > 0$.

There is an origin–destination pair od for which the least cost of travel $\min_{p \in P_{od}} C_p(\mathbf{t}) < C_q(\mathbf{t})$ for some $q \in P_{od}$ that has $t_q > 0$ so the assignment is not in equilibrium. By continuity of the cost functions and convexity of \mathcal{D} , $(\mathbf{s}^*(\mathbf{t}) - \mathbf{t})$ provides a feasible descent direction for the gap. Reassigning some non-zero amount of flow from \mathbf{t} to $\mathbf{s}^*(\mathbf{t})$ will therefore reduce the gap.

3.4 Solution properties: existence and uniqueness

The existence of an equilibrium assignment \mathbf{t}^* follows from the two cases in Section 3.3 and in particular from the descent direction in the second case. According to this, at any assignment $\mathbf{t} \in \mathcal{D}$ either $g(\mathbf{t}) = 0$ so that the flows \mathbf{t} are in equilibrium or the gap is strictly positive and there is a feasible descent direction. A consequence of this is that there exists $\mathbf{t}^* \in \mathcal{D}$ with $g(\mathbf{t}^*) = 0$.

Uniqueness of equilibrium link flows follows from strict monotonicity of the link cost functions. Suppose that $\mathbf{t}^* \in \mathcal{D}$ is an equilibrium assignment, that $\mathbf{t} \in \mathcal{D}$, and let $\mathbf{v}^* = \boldsymbol{\phi} \mathbf{t}^*$ and $\mathbf{v} = \boldsymbol{\phi} \mathbf{t}$. Then by strict monotonicity of costs (3) if $\mathbf{v} \neq \mathbf{v}^*$ then

$$\begin{aligned} [\mathbf{c}(\mathbf{v}^*) - \mathbf{c}(\mathbf{v})]^T \cdot (\mathbf{v}^* - \mathbf{v}) &> 0 \\ \Rightarrow -\mathbf{c}(\mathbf{v})^T \cdot (\mathbf{v}^* - \mathbf{v}) &> \mathbf{c}(\mathbf{v}^*)^T \cdot (\mathbf{v} - \mathbf{v}^*) \geq 0 \end{aligned}$$

(from (7l) because \mathbf{v}^* are equilibrium link flows). We conclude that either $\mathbf{v} = \mathbf{v}^*$, or \mathbf{v} is not an equilibrium and $(\mathbf{t}^* - \mathbf{t})$ is a descent direction for the gap.

A consequence of uniqueness is that if an assignment is in equilibrium, then no different link flows can be. Because the link costs depend on the link flows, the link costs are also uniquely determined at equilibrium as is the total cost of travel. However, because in some cases several distinct route flows lead to the same link flows, no corresponding uniqueness result is available for route flows.

3.5 System optimal assignment

Wardrop's equilibrium principle (W1) represents the behaviour of drivers, and hence the effects of their likely response to changes in demand for travel and the provision of travel facilities. However, according to this, route choice is influenced only by the costs incurred by travellers themselves and not by the resulting system performance. Wardrop recognised that an assignment that minimised the average (and hence, for fixed demand T , the total) cost of travel could differ from equilibrium. He therefore stated the *system optimal principle*:

The average journey time is minimised. (W2)

Evaluating system optimal assignments provides a reference for total cost of travel against which performance of other assignments can be compared. In cases where the difference is small, the assignment can be viewed as making efficient use of the network. By contrast, if the difference is substantial, traffic management measures can be developed to encourage routing that will reduce the total costs. However, in system optimal assignments some travellers will incur costs that are greater than on other routes that are available to them, and these might exceed costs in equilibrium.

The difference in performance between equilibrium and system optimal assignment arises because each traveller imposes cost on others that they do not experience themselves and hence will not consider in making their choices. We analyse this cost externality to calculate a link-based pricing structure that will lead travellers to choose routes that achieve system optimality.

Let $\mathbf{t}^0 \in \mathcal{D}$ be the assignment that minimises the total cost S in the network, i.e. that solves

$$\min_{\mathbf{s} \in \mathcal{D}} \mathbf{C}(\mathbf{s})^T \cdot \mathbf{s}. \quad (9)$$

This differs from the solution to (8) that is used to calculate the descent direction for gap in equilibrium assignment because in that case, the link costs were fixed whereas in (9) they vary with the assignment that is sought.

Provided that the gradients

$$\mathbf{d}(\mathbf{v})^T = \nabla_{\mathbf{v}} S \quad (10)$$

are strictly monotone, (9) is solved by the solution $\mathbf{t}^0 \in \mathcal{D}$ to

$$-\mathbf{d}(\mathbf{v}^0)^T \cdot (\mathbf{v} - \mathbf{v}^0) \leq 0, \quad \forall \mathbf{t} \in \mathcal{D} \quad (11)$$

where $\mathbf{v} = \phi \mathbf{t}$ and $\mathbf{v}^0 = \phi \mathbf{t}^0$. This corresponds to an assignment from which there is no descent direction for the total cost S . Because this formulation corresponds to (71) for equilibrium assignment but with $\mathbf{d}(\cdot)$ in place of the link cost functions $\mathbf{c}(\cdot)$, the corresponding properties of existence, uniqueness and solution approaches obtain.

From (10) and the link-based expression for S :

$$\mathbf{d}(\mathbf{v}) = \mathbf{c}(\mathbf{v}) + \nabla_{\mathbf{v}} \mathbf{c} \mathbf{v}. \quad (12)$$

The second term $\nabla_{\mathbf{v}} \mathbf{c} \mathbf{v}$ on the right-hand side of (12) that augments the costs $\mathbf{c}(\mathbf{v})$ can be interpreted as a measure of the external cost that traffic on each link imposes on other traffic, be it on the same link or on others. Accordingly, if travellers respected this external cost, then the resulting assignment would be system optimal and hence would minimise the total cost of travel in the network. Provided that travellers are homogenous

in their valuation of money, this can be achieved by charging a price $\mathbf{p}^0 = \nabla_{\mathbf{v}} \mathbf{c} \mathbf{v}$ for use of the links of the network so that $\mathbf{d}(\mathbf{v}) = \mathbf{c}(\mathbf{v}) + \mathbf{p}^0(\mathbf{v})$.

If the cost functions $\mathbf{c}(\mathbf{v})$ are separable, according to (12)

$$p_a^0(v_a) = v_a \frac{\partial c_a}{\partial v_a}. \quad (13)$$

If further the cost functions $\mathbf{c}(\mathbf{v})$ have polynomial form, then adding this price to form $\mathbf{d}(\mathbf{v})$ modifies the coefficients. In the particular case of the power function (4)

$$d_a(v_a) = b_a + \alpha(1 + \beta) \left(\frac{v_a}{Q_a} \right)^\beta t_a, \quad \forall a \in L. \quad (14)$$

Just as the cost functions (4) are strictly monotone so are the corresponding price-augmented functions (14).

4 Example calculations

4.1 Introduction

The analysis and methodology described above can be applied to networks of any size. Here we consider assignments to a small network that has a single origin–destination pair and only a few routes. This will illustrate several points concerning route choice that are highlighted by the modelling and analysis presented above.

4.2 Braess's example network

A striking example of the difference between equilibrium and system optimal assignments is provided by Braess's network [9]. This network, which is shown in Figure 2, has a single origin–destination pair with a travel demand $T = 6$ (measured in suitable units). There are three routes between the origin and destination: an upper route passing nodes (a, b, e) , a crossing route (a, b, c, e) , and a lower route (a, c, e) . A crucial feature of this network is that each of the links $[a, b]$ and $[c, e]$ is used by two different routes, one of which uses the central link $[b, c]$.

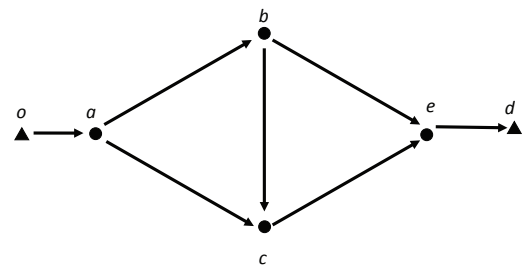


Figure 2: Braess's example network (see, for example, [9]).

The link-route incidence matrix ϕ and the parameters of the affine link cost functions ((4) with $\beta = 1$) are shown in Table 1.

Table 1: Parameters of Braess's example network ($\beta = 1$).

Link $a \setminus$ route p	Incidence matrix ϕ			b_a	$\alpha t_a / Q_a^\beta$
	1	2	3		
$[a, b]$	1	1	0	0	10
$[a, c]$	0	0	1	50	1
$[b, c]$	0	1	0	10	1
$[b, e]$	1	0	0	50	1
$[c, e]$	0	1	1	0	10

4.3 Equilibrium assignment in Braess's network

When the assignment is $\mathbf{t}^* = (2, 2, 2)^T$, the flows and costs are as shown in Table 2. The cost on each route is 92 and there are no unused routes so this is the unique equilibrium assignment according to (2). The total cost of travel is $S^* = 552$.

Table 2: Costs and flows on Braess's network in equilibrium.

	Route p		
	1	2	3
Flow t_p^*	2	2	2
Cost C_p^*	92	92	92

Link a	Flow v_a^*	Cost c_a^*
$[a, b]$	4	40
$[a, c]$	2	52
$[b, c]$	2	12
$[b, e]$	2	52
$[c, e]$	4	40
Total cost S^*	552	

4.4 System optimal assignment in Braess's network

The system optimal assignment that solves (9) and hence minimises the total cost of travel S in the network can be calculated as the equilibrium under the augmented costs $\mathbf{d}(\mathbf{v})$ (12). Because the present cost functions are affine ($\beta = 1$), this cost augmentation doubles the coefficient of link flow, as is shown in Table 3.

Table 3: Parameters of augmented costs functions \mathbf{d} (10) for Braess's example network ($\beta = 1$).

Link a	ba	$\alpha(1 + \beta)t_a/Q_a^\beta$
$[a, b]$	0	20
$[a, c]$	50	2
$[b, c]$	10	2
$[b, e]$	50	2
$[c, e]$	0	20

Calculations based on the assignment $\mathbf{t}^0 = (3, 0, 3)^T$ are given in Table 4 including the associated values $\mathbf{D} = \phi^T \mathbf{d}$ of augmented cost \mathbf{d} accumulated along each route and the values of link (respectively route) cost \mathbf{c} and \mathbf{C} . We see that the augmented cost on routes 1 and 3 (that each carry flow) are equal at 116 and less than the augmented cost on route 2 that carries no flow, which is 130. This verifies the assignment \mathbf{t}^0 as equilibrium in the augmented costs \mathbf{d} and \mathbf{D} , so that it is the unique system optimal assignment that solves the variational inequality (11). These link flows and costs differ from those of equilibrium assignment \mathbf{t}^* as shown in Table 2; in this case the route costs also differ.

In this system optimal assignment \mathbf{t}^0 , the cost of travel on each route that is used is 83, so that the total cost of travel in the network is $S^0 = 498 < 552 = S^*$. That this total cost S^0 does not exceed the total cost S^* at equilibrium is a consequence of the formulation of system optimal assignment as minimising S . In order to achieve this, some travellers must incur lower costs in system optimal assignment than in equilibrium. However, in the present case the cost incurred by each traveller is less than in equilibrium \mathbf{t}^* . Because each of these assignments is uniquely determined, the system optimal assignment cannot be an equilibrium: this is because the crossing route 2 (a, b, c, e) is not used

but has cost 70 that is lower than that of 83 on the routes that are used. This violates the second part of the equilibrium condition (2) because the cost of the routes that are used is not minimal. We therefore see that in this case all travellers will benefit if they all use routes calculated as system optimal instead of changing to a route that currently has lower cost.

Table 4: Costs, prices, augmented costs, and flows on Braess's network with system optimal assignment.

	Route p		
	1	2	3
Flow t_p^0	3	0	3
Route cost C_p^0	83	70	83
Augmented cost D_p^0	116	130	116

Link a	Flow v_a^0	Cost c_a^0	Price p_a^0 (13)	Augmented cost d_a^0
$[a, b]$	3	30	30	60
$[a, c]$	3	53	3	56
$[b, c]$	0	10	0	10
$[b, e]$	3	53	3	56
$[c, e]$	3	30	30	60
Total cost S^0	498	Revenue R^0		198

4.5 Discussion: network design and pricing

Braess's network reveals the cause of the difference between the system optimal and the equilibrium assignments shown in Tables 4 and 2, respectively. This illuminates some issues in the design of transport networks, and the influences that individual choices have on network performance. A good understanding of this is important in designing networks and managing their operation because the resulting performance depends on both the provision and management of the transport system, and on the way in which travellers use it.

In the system optimal assignment \mathbf{t}^0 , the central link $[b, c]$ is not used because the crossing route $[a, b, c, e]$ carries no flow. This route uses both of the links $[a, b]$ and $[c, e]$, each of which has coefficient 10 of flow in the link cost function as shown in Table 1: this is much greater than the coefficients 1 for the other links. This means that each unit of flow on these two links has a large cost externality so that variations in flow on these links will affect the total cost of travel substantially. In the present example, the crossing route uses both of these links so that flow on it will generate a large cost externality twice. By comparison, the other routes 1 and 3 use only one of the links $[a, b]$ and $[c, e]$, so generate the large cost externality only once. The final point is that the cost of the central link $[b, c]$ is low, making it candidate for part of an attractive route: this combination of high cost externality with low cost leads to the difference between the assignments.

We now consider this as an issue in the design of transport networks. In Braess's example network, although the central link carries traffic, its closure and the consequent rerouting of traffic would improve performance with benefit to all travellers. Conversely, if a network corresponded to Braess's example but without the central link $[b, c]$, then the assignment \mathbf{t}^0 would be feasible and an equilibrium. Adding link $[b, c]$ would create an attractive route, but once travellers respond to its presence, the resulting performance would be worsened. Accordingly, there would be no case in reduced transport cost for adding the central link.



If links on the network can be priced, then the matter arises of pricing structures. In Section 3.5, the structure p^0 given by (13) was established as inducing system optimal assignment. The total revenue R collected from link prices p is $R = p^T \cdot v$, which in this case amounts to $R^0 = 198$. The revenue generated can be used, for example, to improve the transport provision. Charging in this way is generally considered preferable to imposing costs through congestion and delay that result in loss of personal time and other resources in a way that is usually unrecoverable and unproductive.

However, this generation of revenue is incidental to the objective of system optimal assignment as can be seen by noting that the optimal price structure is not necessarily unique. In this example network an alternative optimal price structure is 13 (or more) on link $[b, c]$ and 0 on each of the others: this will induce the same system optimal assignment t^0 but will generate no revenue because link $[b, c]$ carries no flow. This shows that the objective of pricing is distinct from that of revenue generation, optimal pricing structures are not necessarily uniquely determined, and that a pricing structure can be optimal without generating revenue.

5 Conclusions

Estimation of traffic flows in network is important in the transport planning process. This involves anticipating the choices of individuals in travel using the various options that are available. The equilibrium model of route choice can be formulated as a variational inequality. Provided that the relationships between costs and flows on the links have appropriate mathematical properties, a unique solution exists in link flows, and link and route costs, though route flows will be indeterminate in some cases. The performance can be calculated and compared with a system optimal value.

This systematic approach shows that increasing traffic capacity in parts of a network can lead to increased costs of travel and hence worsened performance when travellers respond to the new

opportunities. This phenomenon arises when routes are provided that offer relatively low cost for travellers who use them but with relatively high external cost imposed on others in the network. Unless travellers' responses are considered, evaluation of transport provision and traffic management will be unreliable as in some cases this can even reverse the sign of anticipated changes between net benefits and detriments.

REFERENCES

- 1 Peluffo, N. (2015) *Strategic Road Network Statistics*, Department for Transport, London, www.gov.uk/government/uploads/system/uploads/attachment_data/file/448276/strategic-road-network-statistics.pdf.
- 2 Department for Transport (2013) *Action for Roads: A network for the 21st century*, Department for Transport, London, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/212590/action-for-roads.pdf.
- 3 Bar Gera, H. (2010) Traffic assignment by paired alternative segments, *Transp. Res. (Part B: Methodol.)*, vol. 44B, pp. 1022–1046.
- 4 Wardrop, J.G. (1952) Some theoretical aspects of road traffic research, *Proc. Instit. Civil Engin.*, vol. 1, no. 2, pp. 325–362.
- 5 Bureau of Public Roads (1964) *Traffic Assignment Manual*, Department of Commerce, Washington, DC.
- 6 Smith, M.J. (1979) The existence, uniqueness and stability of traffic equilibria, *Transp. Res. (Part B: Methodol.)*, vol. 13, no. 4, pp. 295–304.
- 7 Patriksson, M. (2004) Algorithms for computing traffic equilibria, *Networks Spatial Econ.*, vol. 4, no. 1, pp. 23–38 (see also: www.math.chalmers.se/~mipat/LATEX/NSE.ps).
- 8 Gentile, G. (2014) Local user cost equilibrium: a bush-based algorithm for traffic assignment, *Transportmetrica A: Transp. Sci.*, vol. 10, no. 1, pp. 15–54.
- 9 Murchland, J.D. (1970) Braess's paradox of traffic flow, *Transportmetrica A: Transp. Sci.*, vol. 4, no. 4, pp. 391–394.

Symbols used in text (in order of appearance)

- T vector of travel demands with elements T_{od} from origin o to destination d
 v vector of link flows with v_a the flow on link a , $\forall a \in L$
 L set of links in the network
 ϕ binary link-route incidence matrix:

$$\phi_{ap} = \begin{cases} 1, & \text{if route } p \text{ includes link } a, \\ 0, & \text{otherwise,} \end{cases}$$

$$\forall a \in L, \forall p \in P_{od}, \forall od.$$

- P_{od} set of routes from origin o to destination d , $\forall od$
 t vector of route flows with t_p the flow on route p
 C vector of route costs, with C_p the cost of travel on route $p \in P_{od}$, $\forall od$
 K vector of min. cost of travel, with K_{od} the min. cost between o and d , $\forall od$
 θ binary matrix that identifies the origin–destination pair served by each route:

$$\theta_{p\ od} = \begin{cases} 1, & \text{if } p \in P_{od}, \\ 0, & \text{otherwise,} \end{cases} \quad \forall od.$$

- $c(\cdot)$ vector function of link costs with $c_a(\cdot)$ the cost of travel on link a , $\forall a \in L$
 J Jacobian matrix of the link cost function $J = \nabla_v c$
 $b_a = c_a(0)$ travel cost on link a when $v_a = 0$, $\forall a \in L$
 Q_a reference flow for link a , $\forall a \in L$
 t_a free-flow travel time on link a , $\forall a \in L$
 α, β parameters of the link cost functions: $\alpha > 0, \beta > 0$
 \mathcal{D} set of all feasible route assignments
 $g(t)$ gap function at assignment t , providing a measure of the departure of t from equilibrium
 S total travel cost in network $S = C(t)^T \cdot t = c(v)^T \cdot v$
 $d(v)^T$ grad of S with respect to link flows v : $d = \nabla_v S$
 p vector of link prices, p_a the price on link a , $\forall a \in L$
 R total revenue from link prices p : $R = p^T \cdot v$